DNA Barcoding of Stone Fish *Uranoscopus Oligolepis*: Intra Species Delineation and Hypothetical Protein Analysis

Chandan Tiwary[1], M.A.Badhul Haq*[1], S.Vaitheeswari[1], M.Kalaiselvi[1], Mohammad Nurul Azim Sikder[2] and Wah Wah Min[3]

[1]*Faculty of Marine Sciences, CAS in Marine Biology, Annamalai University, Parangipettai – 608502, India*
[2]*Institute of Marine Sciences and Fisheries, University Of Chittagong, Chittagong-4331, Bangladesh*
[3]*Department of Zoology, University of Yangon, Myanmar.*

*****Corresponding Author: drhaqmarinevirology@gmail.com

Abstract: *The present study addresses this issue by examining the patterning of Cytochrome Oxidase I diversity in the stone fish Uranoscopus oligolepis the structurally diverse group of Family Uranoscopidae. The sequences were analyzed for their species identification using BOLD's identification engine. The COI sequences of U. oligolepis from different geographical regions were extracted from NCBI for intra species variation analysis. All sequences were aligned using Clustal W. The sequences were trimmed using software and phylogenetic tree was constructed with bootstrap test. The results showed that the cytosine content was high (31%). The least molar concentration was observed in guanine (19.5%) and Adenine (19.6%). Thymine was the second predominant in molar concentration next to thymine which is followed by adenine. The G+C content was found to be 49.6% and A+T content was 50.4%. Leucine and Alanine content was high in the amino acid composition. From the study it is assumed that the mitochondrial gene COI can be the potential barcoding region to identify an organism up to the species level.*

Keywords: *COI, intra species, Uranoscopus oligolepis, barcoding, phylogenetic*

INTRODUCTION

DNA barcoding—the sequencing of a short standardized region of DNA—has been proposed as a new tool for animal species identification [1]. The DNA barcode itself consists of a 648 bp region of the cytochrome c oxidase 1 (COI) gene. Additionally to the mitochondrial COI gene, nuclear loci are sometimes also considered to improve assignment performance [2, 3]. This has been shown to provide species level resolution of the vast bulk of species in a wide range of animal taxa, including ants, bats, birds, butterflies, crustaceans, fish, and spiders [4, 5,6,7,8 and 9]. It is based on the postulate that every species will most likely have a unique DNA barcode (indeed there are 4^{650} possible ATGC-combinations compared to an estimated 10 million species remaining to be discovered, Wilson [10].

DNA barcoding has been effectively tested on diverse taxa, from invertebrates [11, 12, 13] to vertebrates [14] allowing the discrimination of different species, often consistently with traditional morphological approaches [15]. The escalating use of DNA barcoding approach in the identification of fish species [16] insisted in raising a new research project with the support from Consortium for the Barcoding of Life (http:// www.barcoding.si.edu/): the Fish Barcode of Life initiative (FISHBOL; http://www.fishbol.org), in which the sequence datas are included into a main unique database called BOLD (Barcode of Life Data System, http:// www.barcodinglife.org/views/login.php; [17, 18]. The cost and time-effectiveness of DNA barcoding enables automated species identification, which is particularly useful in large sampling campaigns. In this way, DNA barcoding could also improve large surveys aiming at unknown species detection and identification of pathogenic species with medical, ecological and agronomical significance [19].

MATERIALS AND METHODS

2.1. Wet lab methodologies

2.1.1. Sample preservation

The *U. oligolepis* fish samples were collected from Parangipettai (South east coast of India) fish landing centre and the tissue samples was excised and cut into small pieces (< 5-7 mm) and preserved in fresh 95% ethanol using 1.5 ml labeled tubes.

2.1.2. DNA extraction

Salting out procedure was adapted to extract DNA from *U. oligolepis* tissues. The preserved tissue in ethanol was washed four to five times with sterile distilled water to get clear of the ethanol content. The ethanol free tissues were transferred in to 1.5 ml tube and grounded in micro pestle with 500µl of solution 1 (500mM Tris-HCL, 20mM EDTA and 2% SDS). After homogenizing the tissues were added with 5µl of Proteinase K (20mg/ml). The tubes were incubated at 55ºC in water bath for 2 hours with occasional mixing by inverting the tubes. Following incubation the samples were chilled on ice for 10 minutes and about 250µl of solution 2 (6M NaCl) was added to it and mixed well by inverting the tubes several times. Tubes were then chilled on ice for 5 minutes. Then the tubes were centrifuged at 8000 rpm for 15 minutes and following centrifugation, 500 µl of clear supernatant was collected in a 1.5 ml tube. Equal

volume of (1ml) of 100% analytical grade ethanol was added to precipitate the DNA. A thin hair like precipitate was observed after addition of ethanol. After 30 minutes the tubes were allowed to spin at 11,000rpm for 5 minutes. The supernatant was removed and partially dried in room temperature. The DNA pellets were washed thrice with 70% cold ethanol. The pellets were suspended in 100 µl of sterile distilled H_2O.

Quantitation of DNA by Spectrophotometric method [20]

- 10µl of DNA solution was diluted with 990µl of TE.
- Mixed well and absorbance at 260nm and 280nm was measured.

The absorbance at 260nm can be used to calculate the concentration of DNA as follows:

Calculations

OD_{260} of 1= 50µg/ml DNA

7Dilution factor =100

Concentration of DNA in a given solution = $\frac{50 \times OD \times \text{Dilution factor}\ \mu g/\mu l}{1 \times 1000}$

2.1.3. PCR amplification

The primer set MAB Fw and MAB Rw designed in the conserved region was used for the amplification of the COI region of the test organisms and the primer sequences are;

MABFw: (5'-TCAACCAACCACAAAGACATTGGCAC-3') and MABRw: (5'-TAGACTTCTGGGTGGCCAAAGAATCA-3'). A 1.0µl of Sample DNA (approximately 100 ng/µl) was added to PCR Mixture containing 100mM Tris HCl (pH 8.3), 500mM KCl (pH 8.3), 2.5µl $MgCl_2$ (25mM), 2.0µl dNTP's (2.5mM), 1.0µl Primer Forward & Reverse (each of 10pm/µl) and 1u /µl of Taq Polymerase (Bioserve Make) & the final volume made to 25 µl with nuclease free water. The thermal profile consisted of 35 cycles at 94 ºC for 50 s, 54 ºC 50 s and 72 ºC for 1 min. The amplicons obtained were about 629 bp long.

2.1.4. DNA sequencing

QIAGEN QIAquick™ kit was used for sequencing reaction. PCRs products were gel purified and directly sequenced using MegaBase sequencer- Bioserve India, Hyderabad. Sequences were checked by eye with Bioedit sequence alignment editor using GenBank sequences as reference sequences and unambiguously aligned using Clustal W.

2.2. Dry lab methodologies

2.2.1. BOLD's identification engine

BOLD (Barcoding of life database) is an online workbench that aids in collection, management, analysis, and use of DNA barcodes. Identification engine is the one of the important components of BOLD database which consists of large volume of barcode sequences for both plants (intranuclear spacer gene) and animals (cytochrome c oxidase subunit gene). BOLD-IDS provide a species identification tool that accepts DNA sequences from the barcode region and returns a taxonomic assignment to the species level when possible. The BOLD identification system (IDS) accepts sequences from the 5' region of the mitochondrial gene cytochrome oxidase subunit I and returns species-level identification when one is possible. Further validation with independent genetic markers will be desirable in some forensic applications. This identification engine was accessible online through http://www.barcodinglife.org/views/ idrequest. php. The sequences were given in FASTA file format in the query box and results were obtained similar to that of BLAST search.

2.2.2. Profiling the barcode region of *Uranoscopus oligolepis*

The molecular weight of the single stranded barcode DNA was calculated as the sum of the monophosphate forms of each deoxyribonucleotide minus one water molecule each. One water (18 Da) was added at the end to represent the 3' hydroxyl at the end of the chain and one more hydrogen atom at the 5' phosphate end. Nucleotide composition summaries and plots were obtained by choosing "Nucleotide Composition" form the "Nucleic Acid" submenu of the "Sequence" menu. Bar plots showed the Molar percent of each residue in the sequence. The degenerate nucleotide designations were added to the plot wherever they are encountered. Any DNA sequence has only A, G, C and T and these were represented by four bars on the graph.

2.2.3. Barcode protein profiling

DNA to Protein: The online software at www.insilico.ehu.es was used to extract hypothesized amino acid sequences from the COI region of *U. oligolepis.* This software allowed modeling and modifications of already existing techniques, as well as new theoretical approaches. Standard genetic code translation was used. DNA sequences were fed in to the query box in FASTA format. Minimum size of protein sequence for Open Reading Frames (ORF) is customizable and they were trimmed to MET-to-Stop.

CLUSTAL W: ClustalW is a general purpose global multiple sequence alignment program for DNA or proteins. It produces biologically meaningful multiple sequence alignments of divergent sequences. It calculates the best match for the selected sequences,

and lines them up so that the identities, similarities and differences can be seen. Evolutionary relationships can be seen via viewing Cladograms or Phylograms.

Phylogenetic tree construction using MEGA: Neighborhood joining (NJ) method of phylogenetic tree construction was preferred for accurate establishment of phylogenetic relationship and to trace out the presence of phylogenetic signals in the DNA sequences [21]. The distance was calculated between every pair of sequences and these were used to construct the phylogenetic tree which guided the final multiple alignment. The scores were calculated from separate pair wise alignments.

MEGA (Molecular Evolutionary Genetic Analysis) version 5: MEGA is an integrated tool for conducting automatic and manual sequence alignment, inferring phylogenetic trees, mining the web base data bases, estimating the rates of molecular evolution, and testing evolutionary hypothesis [22].

Bootstrapping: One of the most commonly used tests of the reliability of an inferred tree is Felsenstein's [23] bootstrap test which is evaluated using [24] bootstrap resampling technique. If there are *m* sequences, each with *n* nucleotides (or codons or amino acids) a phylogenetic tree can be reconstructed using the same tree building method. From each sequence *n* nucleotides were randomly chosen with replacements, giving rise to *m* rows of *n* columns each. These now constitute a new set of sequences.

A tree is then reconstructed with these new sequences using the same tree building method as before. Next the topology of this tree was compared to that of the original tree. Each interior branch of the original tree that was different from the bootstrap tree the sequence it partitions is given a score of 0 all other interior branches was given the value 1 was noted. This procedure of re-sampling the sites and the subsequent tree reconstruction was repeated several hundred times and the percentage of times each interior branch was given a value of 1 was noted. This is known as the bootstrap value.

The multiple aligned sequences from Clustal X were loaded into MEGA through Create New Alignment option in Alignment menu. The sequences were trimmed for their conserved regions and saved in MEGA format for phylogram construction. Bootstrap test for phylogeny was preferred to detect the reliability of each branch in phylogram. As a general rule if the bootstrap value for a given interior branch is 95% or higher than the topology of that branch then the value is considered "correct" [25].

RESULTS

3.1. Quantitation of DNA by electrophoresis

A thick band was seen above the 300kb band of λHind III marker (marker not shown here). This indicates high molecular nature of genomic DNA. In the electropherogram the bands of the size ~700bp (for sample MAB03) was observed against 100bp DNA ladder. There was no overlapping of the bands in the case of test organisms and that way the bands were clear.

3.2. Top 10 Sequences Producing Significant Alignments from NCBI

The sequences were checked for considerable alignments from NCBI. About 10 sequences showed significant alignments of which the maximum identity ranged from 86% to 100%. The maximum score ranged from 1135 to 742. The query coverage was found to be 100%. The summary of the results is depicted in Fig 1. The distance tree comparison of the study organism showed similar evolutionary similarity with *U. oligolepis* (Fig 2).

3.3. BOLD's search

Identification summary (Fig 3) showed the probability of placement (100%) along with taxonomic level and taxon assignment. The distance summary is illustrated in Fig 4. A species level match has been made. This identification is solid unless there is a very closely allied congeneric species that has not yet been analyzed. The bolds search showed top 20 specimen similarity with 88.14% to 100%. (Fig 5). The COI species database tree confirmed that the study organism belongs to the order Carangidae which resembled much similarity with *U. oligolepis* (Fig. 6).

3.4. Accession numbers of sequences closely related to the Test organism used in the analysis & their locations.

The test organisms were reviewed for close relations to the test organism (MAB03 -JX120607) for which the accession numbers were cross checked from the database. The accession numbers were HM422426, GU804916, GU805062 (Ontario) and FJ237963 (China).

3.5. Profiling the barcode region of *Uranoscopus oligolepis*

Nucleotide composition summaries were obtained and shown in Table 1. The table represents the Molar concentration of DNA nucleotides in the COI region *U. oligolepis* sample (MAB03) from Parangipettai waters versus closely related organisms. The results showed that the Cytosine content was high (31%). The least molar concentration was observed in Guanine (19.5%) and Adenine (19.6%). Thymine was the second predominant in molar concentration next to thymine which is followed by adenine. The G+C content was found to be 49.6% and A+T content was 50.4%. Upon comparison with other samples the GC content ranged from 49.6% to 52.3% and the AT content ranged from 47.5 to 50.4%.

3.6. Barcode protein profiling

3.6.1. DNA to Protein

The translation alignment was optional, and amino acids were displayed as a 1-letter amino acids code. Amino acid composition summaries and plots were obtained by choosing "Amino Acid Composition" from the "Protein" submenu of the "Sequence" menu. Bar plots showed the Molar percent of each residue in the sequence (Fig 7). Amino Acid plots and summaries were similar, though residues other than the standard 20 amino acids were ignored. Leucine and Alanine content was high in the amino acid composition. A helical wheel is a type of plot or visual representation used to illustrate the properties of alpha helices in proteins. The sequence of amino acids that make up a helical region of the protein's secondary structure are plotted in a rotating manner where the angle of rotation between consecutive amino acids is 100°, so that the final representation looks down the helical axis. The plot reveals whether hydrophobic amino acids are concentrated on one side of the helix, usually with polar or hydrophilic amino acids on the other (Fig 8)

3.6.2. CLUSTAL W

The similarities between two or more DNA sequences were compared using multiple sequence alignments. The query sequences were posted on the query box in Clustal W from the tools option of EMBL. The results page displays the similarities between the sequences. The similarities in sequences of the study animal with intra species is shown in fig 11. Minimum evolutionary distance of 0.1 in the scale was observed (Fig 9).

DISCUSSION

The efficiency of DNA barcoding has been reported in the detection and description of new cryptic species [26, 27, and 28]. This identification tool can clearly give support to improve classifications and to critically examine the precision of morphological traits commonly used in taxonomy [29]. For a barcoding approach to succeed, within species DNA sequences need to be more similar to one another than those between species and recent studies confirmed that the majority of species examined are well delineated by a tight cluster of very similar sequences [30, 31, 32, 33 and 34]. The methodology requires that intra-species DNA barcode variation is substantially less than interspecies variation, allowing accurate identification of individuals [35]. In the present study *U. oligolepis* from China origin showed more similarity with *U. oligolepis* of Parangipettai waters than with the same species from Ontario. Phylogeographical signals and the arrangement of *U. oligolepis* from China in a separate branch indicate that environmental parameters influence genetic diversity among the same species of organisms.

Freshwater fishes show more population differentiation than marine species, although marine species can show significant differentiation [36] Indeed, several studies have already illustrated the advances provided by the iterative processes between morphological- and DNA barcode-based studies in [37,38 and 39]. Exploring the microscopic eukaryotic life diversity can be achieved by the COI-based barcode [40, 41 and 42]. The profiling study on the barcode regions of *U. oligolepis* revealed that barcode region was rich in cytosine and least in adenine content. The molar concentration of cytosine was found higher when compared to other nucleotides in barcode region of *U. oligolepis* from Parangipettai waters where as the molar concentration of adenine was in the lower side. From the analysis it is assumed that the mitochondrial gene COI can be the potential barcoding region to identify an organism up to the species level. This study clearly revealed that COI could be a barcode sequence distinguishing *U. oligolepis* to its species level both through the phylogram and by search result of barcode of life database.

ACKNOWLEDGMENTS

This work was supported by the SERB, Department and Science and Technology Grant no: SR / FT / LS-125 / 2011. The authors thank the authorities of Annamalai University for their constant support and encouragement.

REFERENCES

[1] P.D.N.Hebert, A.Cywinska, S.L.Ball and J.R.De Waard. Biological identifications through DNA barcodes. Proc. R. Soc. B., 270, 2003, 313–321.

[2] F.Austerlitz, O.David, B.Schaeffer, K.Bleakley, M.Olteanu, R.Leblois, M.Veuille, and C.Laredo. DNA barcode analysis: a comparison of phylogenetic and statistical classification methods. BMC Bioinformatics, BMC Bioinfo.10 (Suppl 14), 2009, S10.

[3] M.Elias, R.I.Hill, K.R.Willmott, K.K.Dasmahapatra, A.V.Brower, J.Mallet and C.D.Jiggins. Limited performance of DNA barcoding in a diverse community of tropical butterflies. Proc. R. Soc. B. 274, 2007, 2881–2889.

[4] P.D.N. Hebert and T.R.Gregory. The promise of DNA barcoding for taxonomy. Syst. Biol. 54, 2005, 852–859.

[5] P.D.N. Hebert, S.Ratnasingham and J.R. DeWaard. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. B .270, 2003: S96–S99.

[6] R.D.H. Barrett and P.D.N Hebert. Identifying spiders through DNA barcodes. Can. J. Zool. 83, 2005,481–491

[7] M.A.Smith, B.L.Fisher and P.D.N.Hebert. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philos Trans R Soc .Lond. B. 360, 2005, 1825–1834.

[8] E.L.Clare, B.K.Lim, M.D.Engstrom, J.L.Eger and P.D.N.Hebert. DNA barcoding of neotropical bats: species identification and discovery within Guyana. Mol Ecol Notes, 7, 2007,184–190.

[9] F.O.Costa, J.R.deWaard, J.Boutillier, S.Ratnasingham, R.T.Dooh, M.Hajibabaei and P.D.N.Hebert. Biological identifications through DNA barcodes: the case of the Crustacea. Can. J. Fish. Aquat. Sci. 64, 2007,272–295.

[10] O.E.Wilson, The encyclopedia of life. Trends. Ecol. Evol.,18:2, 2007, 77–80.

[11] M.Hajibabaei, D.H.Janzen, J.M.Burns, W.Hallwachs and P.D.N.Hebert. DNA barcodes distinguish species of tropical Lepidoptera. Proc. Natl. Acad. Sci. USA.103, 2006, 968–971.

[12] P.D.N.Hebert, E.H.Penton, J.M.Burns, D.H.Janzen and W.Hallwachs. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. USA.101, 2004, 14812–14817.

[13] I.D.Hogg and P.D.N.Hebert. Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes. Can J Zoo., 82, 2004, 749–754.

[14] H.S.Yoo, J.Y.Eah, J.S.Kim, Y.J.Kim, M.S.Min and W.K.Paek. DNA barcoding Korean birds. Mol. Cells., 22, 2006, 323–327.

[15] E.Ferri, M.Barbuto, O.Bain, A.Galimberti, S.Uni and R.Guerriero. Integrated taxonomy: traditional approach and DNA barcoding for the identification of filarioid worms and related parasites (Nematoda). Front Zoo., 6, 2009, 1–25.

[16] E.H.K.Wong and R.H.Hanner. DNA barcoding detects market substitution in North American seafood. Food Res Int., 41, 2008, 828–837.

[17] S.Ratnasingham and P.D.N.Hebert. BOLD: The barcode of life data system. Mol Ecol Notes.7, 2007, 355–367.

[18] M.Barbuto, A.Galimberti, E.Ferri, M.Labra, R.Malandra, P.Galli and M.Casiraghi. DNA barcoding reveals fraudulent substitutions in shark seafood products: The Italian case of ''palombo" (Mustelus spp.). Food Res Int., 43, 2010, 376–381.

[19] Armstrong KF and Ball SL (2005). DNA barcodes for biosecurity: invasive species identification. Phil Trans R Soc B.360: 1813 – 1823.

[20] Badhul Haq MA, Mohammed Azarudeen D, Vignesh R, Ajith Kumar TT and Srinivasan M (2013). Identification and intra species delineation of ornamental silver pompano (*Trachinotus blochi*, Lacepede, 1801) with DNA barcodes**.** International Research Journal of Biochemistry and Bioinformatics.Vol. 3(1):26-36.

[21] Nei M, Kumar S and Takahashi K (1998). The optimization principle in phylogenetic analysis tends to give incorrect topologies when the number of nucleotides or amino acids used is small. Proc Nat Acad Sci USA. 95:12390–12397.

[22] Tamura K, Dudley J, Nei M and Kumar S (2007). MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol Bio Evol. 24:1596-1599.

[23] Felsenstein J (1985). Confidence limits on phylogenies: An approach using the bootstrap. Evolution. 39: 783–791.

[24] Efron B (1982). The Jackknife, the Bootstrap and Other Resampling Plans. CBMS-NSF Regional Conference Series in Applied Mathematics, Monograph 38, SIAM, Philadelphia.

[25] Nei M and Kumar S (2000).Molecular evolution and Phylogenetics. Oxford University press., New York , pp. 333.

[26] Anker A, Hurt C and Knowlton N (2007). Revision of the *Alpheus nuttingi* (Schmitt) species complex (Crustacea:Decapoda:Alpheidae), with description of a new species from the tropical eastern Pacific. Zootaxa. 1577: 41–60.

[27] Bucklin A, Wiebe PH, Smolenack SB, Copley NJ, Beaudet JG, Bonner KG, Farber-Lorda J and Pierson JJ (2007). DNA barcodes for species identification of euphausiids (Euphausiacea, Crustacea). J. Plankton. Res. 29: 483–493.

[28] Tavares ES and Baker AJ (2008). Single mitochondrial gene barcodes reliably identify sister-species in diverse clades of birds. BMC Evol. Biol. 8: 81.

[29] Lise Frezala and Raphael Leblois (2008). Four years of DNA barcoding: Current advances and prospects. Infection .Genetics. Evol. 8:727–736.

[30] Robins JH, Hingston M, Matisoo-Smith E and Ross HA (2008). Identifying Rattus species using mitochondrial DNA. Mol. Ecol. Notes. 7: 717–729.

[31] Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM and Hebert PDN (2007). Comprehensive DNA barcode coverage of North American birds. Mol.Ecol.Notes 7: 535–543.

[32] Hubert N, Hanner R, Holm E, Mandrak N, Taylor E, Burridge M, Watkinson DA, Dumont P, Curry A, Bentze P, Zhang J, April J and Bernatchez L (2008) . Identifying Canadian freshwater fishes through DNA barcodes. PLoS ONE. 3: e2490.

[33] Foottit RG, Maw HEL, Havill NP, Ahern RG and Montgomery ME (2009). DNA barcodes to identify species and explore diversity in the Adelgidae (Insecta: Hemiptera: Aphidoidea). Mol. Ecol. Res. 9: 188–195.

[34] Sheffield CS, Hebert PDN, Kevan PG and Packer L (2009). DNA barcoding a regional bee (Hymenoptera: Apoidea) fauna and its potential for ecological studies. Mol. Ecol. Res. 9: 196–207.

[35] Ward RD, Woodwark M and Skibinski DOF (1994). A comparison of genetic diversity levels in marine, freshwater and anadromous fish. J. Fish. Biol. 44:213–232.

[36] Page TJ, Choy SC and Hughes JM (2005). The taxonomic feedback loop: symbiosis of morphology and molecules. Biol Lett. 1: 139–142.

[37] Carlini DB, Kunkle LK and Vecchione M (2006). A molecular systematic evaluation of the squid genus Illex (Cephalopoda: Ommastrephidae) in the North Atlantic Ocean and Mediterranean Sea. Mol Phylo. Evol. 41: 496–502.

[38] Van Velzen R, Freek T, Bakker L and Van Loon JJA (2007). DNA barcoding reveals hidden species diversity in Cymothoe (Nymphalidae). Proc.Neth. Entomol. Soc. Meet. 18: 95–103.

[39] Blaxter M, Mann J, Chapman T, Thomas F, Whitton C, Floyd R and Abebe E (2005). Defining operational taxonomic units using DNA barcode data. Philos. Trans. R. Soc. Lond. B. 360: 1935–1943.

[40] Chantangsi C, Lynn DH, Brandl MT, Cole JC, Hetrick N and Ikonomi P (2007). Barcoding ciliates: a comprehensive study of 75 isolates of the genus Tetrahymena. Int. J. Syst. Evol. Microbiol.57: 2412–2425.

[41] Evans KM, Wortley AH and Mann DG (2007). An assessment of potential diatom ''barcode'' genes (cox1, rbcL, 18S and ITS rDNA) and their effectiveness in determining relationships in Sellaphora (Bacillariophyta). Protist. 158: 349–364.

[42] Robert D. Ward, Filipe O. Costa, Bronwyn H. Holmes and Dirk Steinke (2008). DNA barcoding of shared fish species from the North Atlantic and Australasia: minimal divergencefor most taxa, but *Zeus faber* and *Lepidopuscaudatus* each probably constitute two species. Aquat. Biol. 3: 71–78.

Sequences producing significant alignments:

Accession	Description	Max score	Total score	Query coverage	E value	Max ident
FJ237963.1	Uranoscopus oligolepis voucher MBCSC:HN SY08334 cytochrome ox	1135	1135	100%	0.0	100%
FJ237961.1	Uranoscopus oligolepis voucher MBCSC:HN SY08506 cytochrome ox	1130	1130	100%	0.0	99%
FJ237962.1	Uranoscopus oligolepis voucher MBCSC:HN SY08335 cytochrome ox	1130	1130	100%	0.0	99%
HM422426.1	Perciformes sp. BOLD:AAJ8254 voucher BW-A7438 cytochrome oxic	908	908	99%	0.0	92%
GU805069.1	Uranoscopus archionema voucher ADC09_230.2#2 cytochrome oxid	848	848	99%	0.0	90%
GU804916.1	Uranoscopus archionema voucher ADC09_230.2#5 cytochrome oxid	848	848	99%	0.0	90%
GU805062.1	Uranoscopus archionema voucher ADC09_230.2#1 cytochrome oxid	843	843	99%	0.0	90%
GU804944.1	Uranoscopus archionema voucher ADC09_230.2 #4 cytochrome oxi	843	843	99%	0.0	90%
HQ920475.1	Perciformes sp. BOLD:AAG7346 voucher I.44773-019 cytochrome o	758	758	99%	0.0	87%
GU674178.1	Perciformes sp. BOLD:AAD7191 voucher BW-A6917 cytochrome oxi	742	742	99%	0.0	86%

Figure 1. Top 10 Sequences producing significant alignments from NCBI

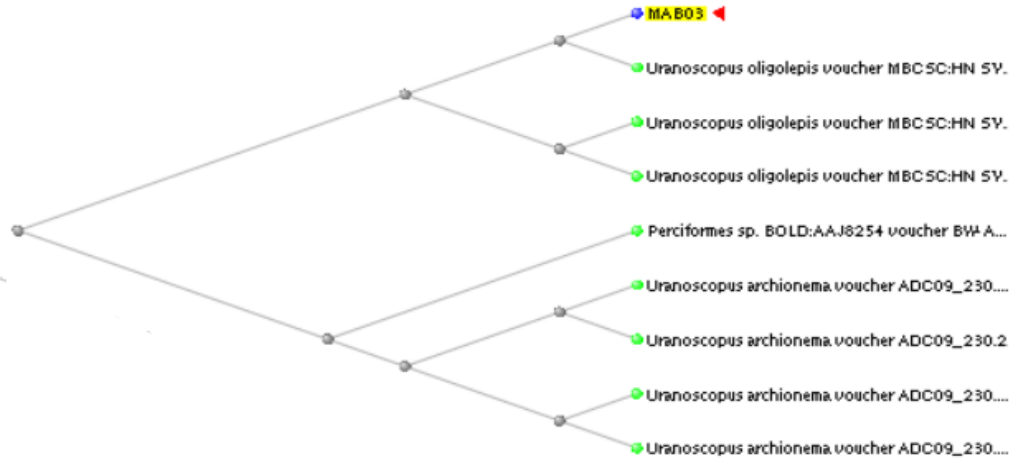

Figure 2. Distance tree of the study animal extracted from NCBI

Search Request:

Type : COI SPECIES DATABASE

Search Result:

Identification Summary :

Taxonomic Level	Taxon Assignment	Probability of Placement (%)
phylum	Chordata	100
class	Actinopterygii	100
order	Perciformes	100
family	Uranoscopidae	100
genus	Uranoscopus	100

A species level match could not be made, the queried specimen is likely to be one of the following :

-Uranoscopus oligolepis
-Uranoscopus sp.

Figure 3. Results of the identification summary from BOLD search

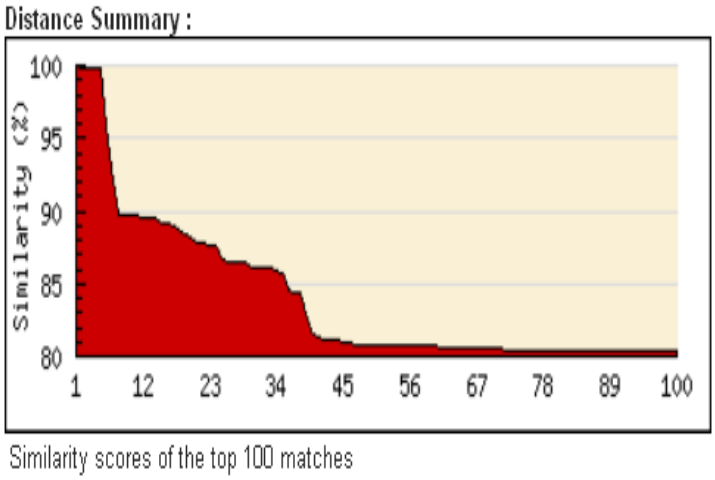

Figure 4. Results of the distance summary

TOP 20 Matches :

Phylum	Class	Order	Family	Genus	Species	Specimen Similarity (%)
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*oligolepis*	100
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*oligolepis*	99.84
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*oligolepis*	99.84
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*sp.*	99.84
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*oligolepis*	99.83
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*sp.*	95.83
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*kaianus*	91.99
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*n.spscalynape*	89.74
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*cognatus*	89.74
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*archionema*	89.74
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*archionema*	89.74
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*archionema*	89.58
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*archionema*	89.58
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*archionema*	89.58
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*japonicus*	89.08
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*japonicus*	89.05
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*japonicus*	88.94
Chordata	Actinopterygii	Pleuronectiformes	Cynoglossidae	Cynoglossus	*arel*	88.7
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*bicinctus*	88.32
Chordata	Actinopterygii	Perciformes	Uranoscopidae	Uranoscopus	*bicinctus*	88.14

Figure 5. Results of top 20 specimen similarity with 88.14% to 100%

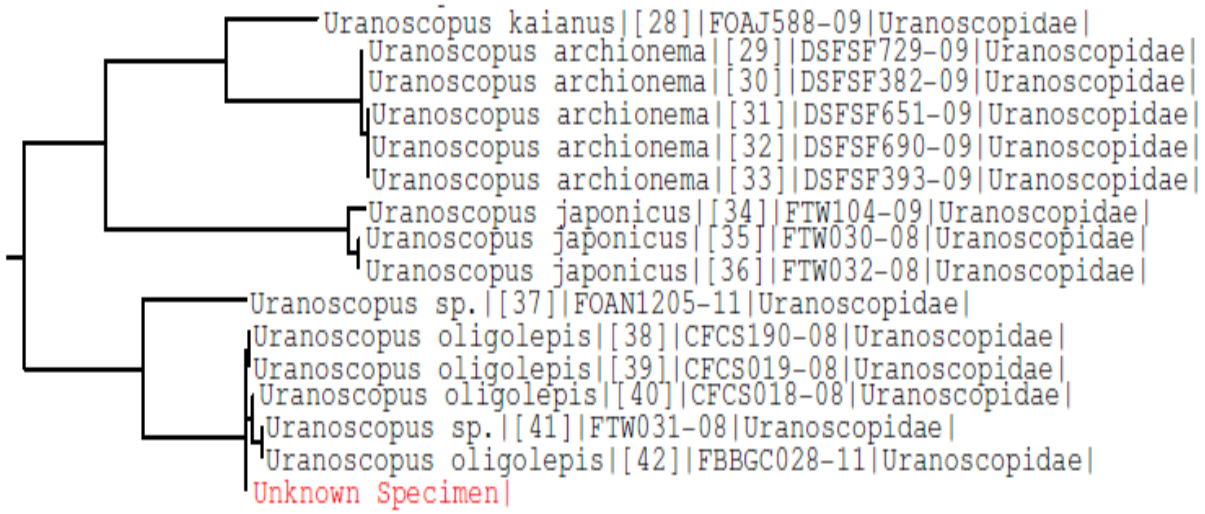

Figure 6. COI species database tree

```
                    250        260        270        280        300
                    ....|....| ....|....| ....|....| ....|....| ....|....|
MAB03       236 GAACAACATG AGCTTCTGGC TCCTACCCCC CTCTCTTGTC CTGCTACTCG CATCTTCCGG
FJ237963    241 GAACAACATG AGCTTCTGGC TCCTACCCCC CTCTCTTGTC CTGCTACTCG CATCTTCCGG
HM422426    241 AAATAACATG AGCTTTTGAC TCCTGCCCCC GTCTCTTGTA CTCCTCCTCG CATCTTCCGG
GU804916    241 AAATAACATG AGCTTTTGAC TCCTGCCTCC GTCCCTTCTC CTCCTCCTCG CATCTTCCGG
GU805062    241 AAATAACATG AGCTTTTGAC TCCTGCCTCC GTCCCTTCTC CTCCTCCTCG CATCTTCCGG

                    310        320        330        340        350        360
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
MAB03       296 AGTTGAGGCC GGAGTCGGCA CCGGTTGAAC TGTTTACCCA CCCTTGGCGG GCAACCTCGC
FJ237963    301 AGTTGAGGCC GGAGTCGGCA CCGGTTGAAC TGTTTACCCA CCCTTGGCGG GCAACCTCGC
HM422426    301 GGTCGAAGCC GGAGTCGGCA CCGGTTGAAC TGTTTACCCG CCCTTGGCGG GCAACCTCGC
GU804916    301 GGTCGAGGCC GGGGCCGGCA CCGGTTGGAC TGTTTACCCG CCCTTGGCGG GGAATCTCGC
GU805062    301 GGTCGAGGCC GGGGCCGGCA CCGGTTGGAC TGTTTACCCG CCCTTGGCGG GGAATCTCGC

                    370        380        390        400        410        420
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
MAB03       356 CCACGCAGGG GCATCCGTGG ACCTTGCTAT TTTCTCCCTT CACCTAGCAG GAATTTCCTC
FJ237963    361 CCACGCAGGG GCATCCGTGG ACCTTGCTAT TTTCTCCCTT CACCTAGCAG GAATTTCCTC
HM422426    361 CCACGCAGGT GCATCCGTAG ACCTTGCTAT TTTCTCCCTT CACCTGGCGG GGATTTCCTC
GU804916    361 CCACGCAGGT GCATCCGTAG ACCTTGCTAT TTTCTCCCTT CACCTCGCGG GGATTTCCTC
GU805062    361 CCACGCAGGT GCATCCGTAG ACCTTGCTAT TTTCTCCCTT CACCTCGCGG GGATTTCCTC

                    430        440        450        460        470        480
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
MAB03       416 TATTTTAGGG GCTATTAATT TTATTACTAC TATTATTAAT ATAAAACCAC CCGGCACCTC
FJ237963    421 TATTTTAGGG GCTATTAATT TTATTACTAC TATTATTAAT ATAAAACCAC CCGGCACCTC
HM422426    421 TATTCTAGGG GCCATTAACT TTATTACTAC TATTATTAAT ATAAAACCGC CCGGCACCTC
GU804916    421 AATTCTGGGG GCCATTAACT TTATCACTAC TATTGTTAAT ATAAAACCGC CCGGCACCTC
GU805062    421 AATTCTGGGG GCCATTAACT TTATCACTAC TATTGTTAAT ATAAAACCGC CCGGCACCTC

                     10         20         30         40         50         60
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
MAB03         1     -----ACTTG GTATTTGGTG CTTGAGCAGC AATAGTAGGA ACAGCCCTGA GCCTACTTAT
FJ237963      1 CCTTTACTTG GTATTTGGTG CTTGAGCAGC AATAGTAGGA ACAGCCCTGA GCCTACTTAT
HM422426      1 CCTTTATCTG GTGTTTGGTG CATGAGCTGC GATAGTGGGA ACAGCCCTGA GCCTACTAAT
GU804916      1 CCTTTATCTG GTGTTTGGTG CATGAGCTGC GATAGTGGGA ACAGCCCTGA GCCTACTAAT
GU805062      1 CCTTTATCTG GTGTTTGGTG CATGAGCTGC GATAGTGGGA ACAGCCCTGA GCCTACTAAT

                     70         80         90        100        110        120
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
MAB03        56 CCGAGCCGAG CTTAATCAGC CCGGCGCCCT ACTCGGGGAC GATCAAATCT ATAATGTGAT
FJ237963     61 CCGAGCCGAG CTTAATCAGC CCGGCGCCCT ACTCGGGGAC GATCAAATCT ATAATGTGAT
HM422426     61 CCGAGCCGAG CTTAGTCAGC CCGGTGCCCT ACTCGGGGAC GATCAAATCT ATAATGTGAT
GU804916     61 CCGAGCCGAG CTTAGTCAGC CCGGTGCCCT ACTCGGGGAC GATCAAATCT ATAATGTGAT
GU805062     61 CCGAGCCGAG CTTAGTCAGC CCGGTGCCCT ACTCGGGGAC GATCAAATCT ATAATGTGAT

                    130        140        150        160        170        180
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
MAB03       116 TGTTACAGCA CATGCTTTTG TAATAATCTT TTTTATGGTT ATGCCAGTAA TAATTGGAGG
FJ237963    121 TGTTACAGCA CATGCTTTTG TAATAATCTT TTTTATGGTT ATGCCAGTAA TAATTGGAGG
HM422426    121 TGTTACGGCA CATGCCTTCG TAATAATCTT TTTTATAGTT ATGCCAGTAA TAATTGGAGG
GU804916    121 TGTTACGGCG CATGCCTTCG TAATAATCTT TTTTATAGTC ATGCCAGTAA TAATTGGAGG
GU805062    121 TGTTACGGCG CATGCCTTCG TAATAATCTT TTTTATAGTC ATGCCAGTAA TAATTGGAGG

                    190        200        210        220        230        240
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
MAB03       176 CTTTGGTAAC TGATTAGTCC CACTAATAAT TGGCGCCCCC GATATAGCAT TCCCCCGAAT
FJ237963    181 CTTTGGTAAC TGATTAGTCC CACTAATAAT TGGCGCCCCC GATATAGCAT TCCCCCGAAT
HM422426    181 CTTTGGTAAC TGATTGGTCC CATTAATAAT TGGCGCCCCC GACATAGCAT TCCCCCGAAT
GU804916    181 CTTTGGTAAC TGATTAGTCC CATTAATAAT TGGCGCCCCC GACATAGCAT TCCCTCGAAT
GU805062    181 CTTTGGTAAC TGATTAGTCC CATTAATAAT TGGCGCCCCC GACATAGCAT TCCCTCGAAT
```

```
                    490         500         510         520         530         540
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
MAB03          476  CCAGTACCAA GTCCCCTTGT TTGTCTGGGC TGTCTTTGTC ACCGCCGTCC TTCTTCTGCT
FJ237963       481  CCAGTACCAA GTCCCCTTGT TTGTCTGGGC TGTCTTTGTC ACCGCCGTCC TTCTTCTGCT
HM422426       481  TCAGTACCAA GTCCCCTTAT TCGTCTGGGC TGTCTTTGTC ACCGCCGTCC TTCTACTACT
GU804916       481  TCAGTACCAG GTCCCCCTAT TCGTCTGGGC TGTCTTTGTC ACCGCCGTCC TTCTACTGCT
GU805062       481  TCAGTACCAG GTCCCCCTAT TCGTCTGGGC TGTCTTTGTC ACCGCCGTCC TTCTACTGCT

                    550         560         570         580         590         600
                    ....|....| ....|....| ....|....| ....|....| ....|....| ....|....|
MAB03          536  CTCCTTACCC GTTTTAGCCG CTGCTATCAC GATGCTCCTA ACAGACCGTA ACCTAAATAC
FJ237963       541  CTCCTTACCC GTTTTAGCCG CTGCTATCAC GATGCTCCTA ACAGACCGTA ACCTAAATAC
HM422426       541  TTCTTTACCG GTCTTAGCCG CTGCTATCAC AATACTCCTA ACAGACCGTA ACCTAAATAC
GU804916       541  TTCCTTACCC GTTTTAGCCG CTGCTATCAC AATACTCCTA ACAGACCGTA ATCTAAACAC
GU805062       541  TTCCTTACCC GTTTTAGCCG CTGCTATCAC AATACTTCTA ACAGACCGTA ATCTAAACAC

                    610         620         630         640         650
                    ....|....| ....|....| ....|....| ....|....| ....|....| ..
MAB03          596  CACCTTCTTT GACCCCGCTG GGGGAGGAGA CCCA------ ---------- --
FJ237963       601  CACCTTCTTT GACCCCGCTG GGGGAGGAGA CCCAATTCTC TATCAACACC TC
HM422426       601  CACCTTCTTT GACCCCGCCG GGGGAGGAGA CCCAATCCTC TACCAGCACC TC
GU804916       601  CACCTTCTTT GACCCCGCCG GGGGAGGGGA CCCGATCCTC TACCAGCACC TC
GU805062       601  CACCTTCTTT GACCCCGCCG GGGGAGGGGA CCCGATCCTC TACCAGCACC TC
```

Figure 7. Clustal W alignments for the test organism versus closely related specimens

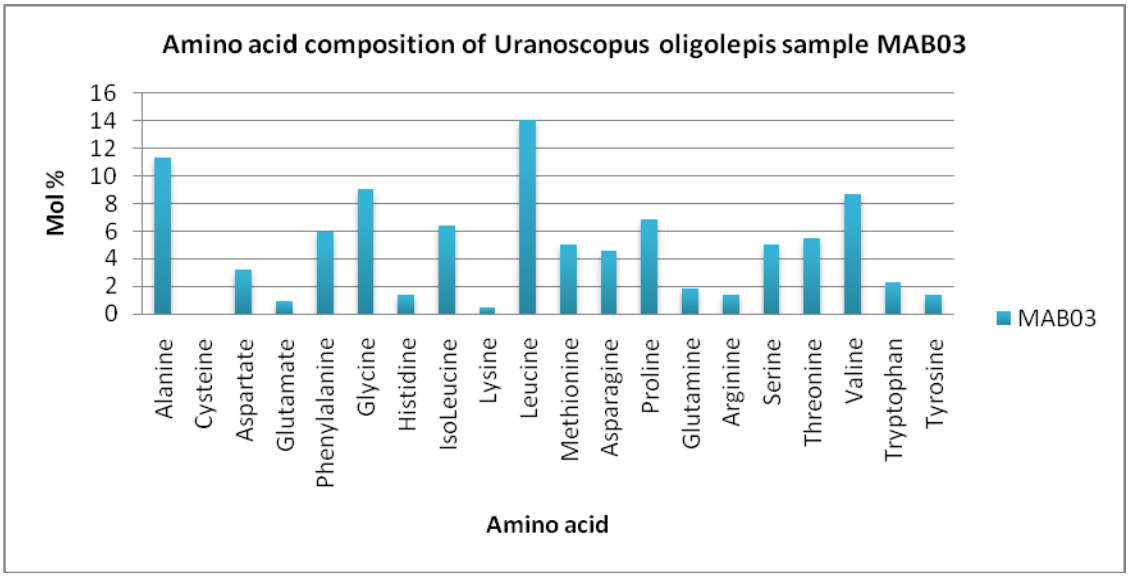

Figure 8. Graph showing hypothetical barcode profiling of *T. blochi* of Parangipettai waters

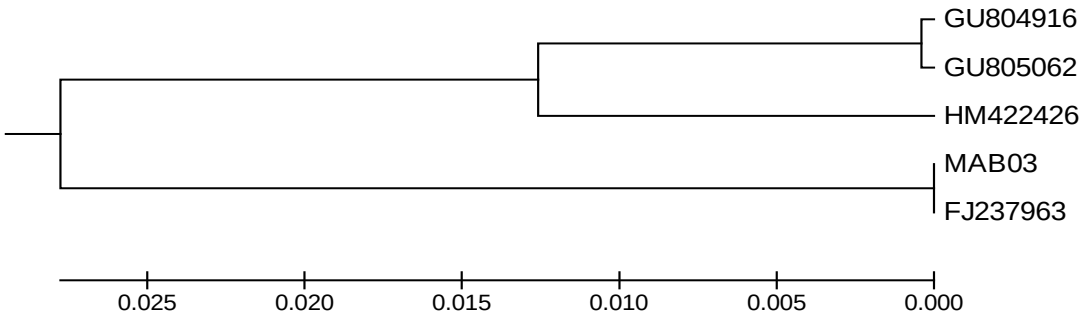

Figure 9. Evolutionary relationships of taxa of study animal (MAB03 - JX120607) by UPGMA Method

Table 1. The Molar concentration of DNA nucleotides in the COI region *U. oligolepis* sample from Parangipettai waters versus closely related specimens

Accession ID	**Base pair length**	**G+C content (%)**	**A+T content (%)**	**Nucleotide Number and Mol%**			
				A	T	G	C
JX120607 (MAB03)	629	49.6%	50.4%	134 21.3%	183 29.1%	128 20.3%	184 29.3%
FJ237963	652	49.2%	50.8%	139 21.3%	192 29.4%	128 19.6%	193 29.6%
HM422426	652	50.6%	49.4%	136 20.9%	186 28.5%	133 20.4%	197 30.2%
GU804916	652	52.5%	47.5%	127 19.5%	183 28.1%	140 21.5%	202 31.0%
GU805062	652	52.3%	47.7%	127 19.5%	184 28.2%	140 21.5%	201 30.8%